

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 9 月 10 日 (10.09.2020)



(10) 国际公布号
WO 2020/177214 A1

- (51) 国际专利分类号:
G06K 9/62 (2006.01) **G06N 3/08** (2006.01)
G06N 3/04 (2006.01)
- (21) 国际申请号: PCT/CN2019/085962
- (22) 国际申请日: 2019 年 5 月 8 日 (08.05.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910172420.0 2019 年 3 月 7 日 (07.03.2019) CN
- (71) 申请人: 北京大学深圳研究生院 (PEKING UNIVERSITY SHENZHEN GRADUATE SCHOOL) [CN/CN]; 中国广东省深圳市南山

区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。

(72) 发明人: 王文敏 (WANG, Wenmin); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。 李炜棉 (LI, Weimian); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。 黄钟毅 (HUANG, Zhongyi); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。

(74) 代理人: 北京万象新悦知识产权代理有限公司 (BEIJING WANXIANGXINYUE INTELLECTUAL PROPERTY OFFICE); 中国北京市海淀区颐和园路 1 号北京资源燕园宾馆六层 1625 号, Beijing 100080 (CN)。

(54) Title: DOUBLE-STREAM VIDEO GENERATION METHOD BASED ON DIFFERENT FEATURE SPACES OF TEXT

(54) 发明名称: 一种基于文本不同特征空间的双流式视频生成方法

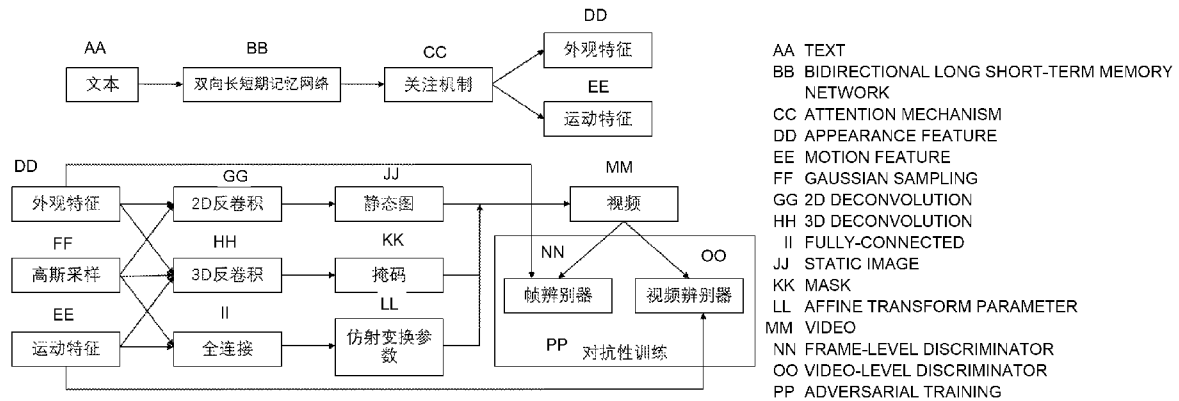


图 1

(57) Abstract: A double-stream video generation method based on different feature spaces of a text, comprising: a text feature extraction process, a double-stream video generation process, and an adversarial training process. In the present invention, according to an object appearance feature space and a corresponding motion feature space thereof described in the text, by separating a space feature and a temporal feature comprised in the text, modeling these features in a double-stream manner respectively, and optimizing a generated result in an adversarial training manner to enable video fragments to be closer to reality, intelligent generation of the video fragments conforming to the text content is realized. The method has wide market demands and application prospects in the aspects of material accumulation and data set automatic generation.

(57) 摘要: 一种基于文本不同特征空间的双流式视频生成方法, 包括: 文本特征提取过程、双流式视频生成过程和对抗性训练过程。本发明根据文本中所描述的物体外观特征空间与其相应的运动特征空间, 通过分离文本中所包含的空间特征和时序特征, 并采用双流的方式分别对这些特征建模, 利用对抗训练的方式来优化生成结果, 使其逐渐趋于真实, 从而实现智能化生成符合文本内容的视频片段, 在素材积累和数据集自动生成方面具有广泛的市场需求和应用前景。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

一种基于文本不同特征空间的双流式视频生成方法

技术领域

本发明涉及模式识别、自然语言处理、计算机视觉等技术领域，尤其涉及一种基于文本不同特征空间的视频生成方法，根据文本中所描述的物体外观特征空间与其相应的运动特征空间，智能化生成符合文本内容的视频片段。

背景技术

近年来，随着计算机视觉领域的飞速发展和生成对抗网络的提出，图像生成的研究受到了越来越广泛的关注，其在素材积累，数据集自动生成方面有非常积极意义。视频相比于图像它更加生动，生成难度也更大，因此对于视频生成方面的探索更加有研究意义。同时，如果与大多数的图像生成方法一样随机地生成视频并不具有太多实用价值，用户更多地会想要基于某种给定信息来生成，比如，用户输入“一个人在走路”，期望得到与所输入文本相匹配的视频片段，而不是一些随机的、无意义的视频。如此看来，传统的生成方法已经不能满足用户对于生成结果指向性的、匹配性的要求。基于文本的视频生成面临的挑战主要是如何有效地理解文本内容（“什么物体在进行着怎样的运动”），并且生成结果要保证其每一帧的生成质量和整个视频在时序上动作的连贯性。

现有的方法大多对整个文本进行统一地处理，对提取到的特征直接生成整个视频片段，但是，这样的处理高估了单一模型的学习能力，不仅需要学习空间特征（外观信息），还要学习时序特征（运动信息），使得模型对于两种特征都不能充分地理解，难以有效地进行高质量的生成。

发明内容

本发明提供了一种基于文本不同特征空间的双流式视频生成方法，根据文本中所描述的物体外观特征空间与其相应的运动特征空间，通过分离文本中所包含的空间特征和时序特征，并采用双流的方式分别对这些特征建模，利用对抗训练的方式来指导生成过程和优化生成结果，使其逐渐趋于真实，从而实现智能化生成符合文本内容的视频片段。本发明还通过在简单物体运动视频数据库中验证了本发明方法的有效性。

本发明的技术方案是：

一种基于文本不同特征空间的双流视频生成方法，通过分离文本中所包含的空间特征和时序特征，并采用双流的方式分别对这些特征建模，并利用对抗训练的方式进行训练，实现智能化生成符合文本内容的真实高的视频片段；包括：文本特征提取过程、双流式视频生成过程和对抗性训练过程。

1) 文本特征提取过程:

本发明解析了输入的文本内容,对得到的特征信息进行了分离,分别提取了外观特征空间与运动特征空间的信息。

需要注意的是,为了确保输入后续视频生成过程的文本特征是有意义的,文本特征提取过程要经过提前训练,并在后续其他模块的训练过程中保持不变。

文本特征提取过程具体包括如下步骤:

11) 提取得到文本的外观特征 T_a 和运动特征 T_m :

对输入的文本,采用双向的长短期记忆网络(bidirectional long short-term memory, LSTM)实现对文本特征的映射,同时引入了注意力关注机制(attention mechanism)结构进行外观特征和运动特征的分离。

具体操作如下:

111) 双向长短期记忆网络分为上下两层,分别对文本做一个正序和逆序的解析,可以得到更全面的理解。文本中的每一个词依次输入长短期记忆网络的每个时间节点(time step),每个时间节点处的维度为256维,提取到的特征分别表示为正序文本特征 $T_x = \{t_{x1}, t_{x2}, \dots, t_{xn}\}$ 和逆序文本特征 $T_y = \{t_{y1}, t_{y2}, \dots, t_{yn}\}$, n表示文本的长度,然后通过取均值的方式得到需要的文本特征 $T = \{t_1, t_2, \dots, t_n\}$,表示为式1:

$$T = (T_x + T_y) / 2 \quad (\text{式 } 1)$$

其中 $T_x + T_y$ 采用的是向量加法。

112) 在注意关注机制结构中,文本特征 T 经过两组并行的全连接网络(fully-connected),全连接网络的最后一层使用 softmax 激活函数,得到两组权重 $w_a = \{w_{a1}, w_{a2}, \dots, w_{an}\}$ 和 $w_b = \{w_{b1}, w_{b2}, \dots, w_{bn}\}$,并且满足 $\sum_{i=1}^n w_{ai} = 1$ 和 $\sum_{i=1}^n w_{bi} = 1$,分别用于提取外观特征和运动特征。以外观特征的提取为例,其计算过程可以表示为式2:

$$T_a = \sum_{i=1}^n t_i * w_{ai} \quad (\text{式 } 2)$$

权重中每个值 w_{ai} 代表了对文本中不同位置特征的关心程度,值越大代表其对于外观特征的描述更多,通过加权求和的方式可得到“关注”后的特征。同理,文本特征中包含的运动特征 T_m 也可以采用这种方式来进行提取。

12) 对提取的外观特征 T_a 和运动特征 T_m ,需要在数据集中选择正负样本,同时引入三元组损失(triplet loss),实现文本特征准确的分离。

具体操作如下：

121) 在训练提取外观特征过程中，定义三元组 $\Phi^a = (T_a, G^+, G^-)$ ，其中 T_a 为参照用外观特征， G^+ 为正样本，是包含符合文本描述的物体的视频帧， G^- 为负样本，是包含文本描述不同的物体的视频帧。此时关注的是外观特征，只需要物体类型符合文本描述即可，使用的仅是视频中的一帧。对选取的视频帧，使用 2D 卷积神经网络 (2D convolution neural network) 来提取其空间特征 (spatial feature)，并在最后利用一个单层全连接网络结构将其映射到 T_a 的同构空间，为后续的图文匹配做准备。使用 L2 范数计算的不同模态数据间距离，表示为式 3：

$$l_2^a(T_a, G^+) = \|T_a - G^+\|_2^2 \quad (\text{式 } 3)$$

其中 $l_2^a(T_a, G^+)$ 表示同构空间内图像与文本之间的欧式距离。

以外观特征 T_a 为参照物，引入三元组约束 (triplet constraint) 调整文本分布，即拉近匹配的文本图像对 (T_a, G^+) 的距离，同时增加不匹配的文本图像对 (T_a, G^-) 的距离，确保分离出的外观特征是准确的、有意义的。外观特征的三元组损失可以表示为式 4：

$$L_{tri}^a(\Phi^a) = \max(0, \alpha - l_2^a(T_a, G^+) + l_2^a(T_a, G^-)) \quad (\text{式 } 4)$$

其中 α 为超参数，表示文本特征与正负样本间距离之差的最大值。

122) 相似地，在训练提取动作特征过程中，选取包含符合文本描述的运动的视频 V^+ ，此时关注的是运动特征，需要使用整个视频，但视频中包含的物体类型可以忽略。对选取的视频帧，使用 3D 卷积神经网络 (3D convolution neural network) 来提取其时序特征 (temporal feature)，并在最后利用一个单层全连接网络结构将其映射到 T_m 的同构空间。

同样的，通过构建三元组 $\Phi^m = (T_m, V^+, V^-)$ 和采用式 4 所示的损失函数计算 $L_{tri}^m(\Phi^m)$ ，拉近匹配的文本视频对 (T_m, V^+) 的距离，增加不匹配的文本视频对 (T_m, V^-) 的距离。

文本特征提取过程的目标函数 L 由外观特征的三元组损失 $L_{tri}^a(\Phi^a)$ 与运动特征的三元组损失 $L_{tri}^m(\Phi^m)$ 组成，可以表示为式 5：

$$L = L_{tri}^a(\Phi^a) + L_{tri}^m(\Phi^m) \quad (\text{式 } 5)$$

2) 双流式视频生成器的构建过程

本发明引入双流式视频生成器来生成视频，先利用提取的外观特征 T_a 生成一张静态图 (static image)，再将学习到的运动特征 T_m 作用于这张图上来衍生出后续的视频帧，即以“先静后动”的方式来生成视频。

执行如下操作：

21) 对外观特征 T_a ，先利用一个单层全连接网络结构将其映射到 $4*4*1024$ 维度，并维度

重组为 (4, 4, 1024), 再通过 2D 反卷积神经网络 (2D deconvolution neural network) 来生成静态图F。其中, 2D 反卷积神经网络由 5 层反卷积层组成, 每层反卷积层的感受野大小为 $4 * 4$, 步长为 2, 最终输出的图像大小为 $64 * 64$ 。同时为了对同样的文本生成更多样性的视频, 引入一个随机变量 Z, 该随机变量从高斯分布采样, 使得生成的静态图中的物体可以有略微不同的外观和不同的初始位置。

22)对运动特征 T_m , 采取的方式是利用其来学习仿射变换 (affine transform) 的参数, 单一的仿射变换有六个参数, 表示为矩阵 $A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & 1 \end{bmatrix}$, 其变换过程可以表示为式 6:

$$\begin{bmatrix} x'_k \\ y'_k \\ 1 \end{bmatrix} = A \begin{bmatrix} x_k \\ y_k \\ 1 \end{bmatrix} \quad (\text{式 6})$$

其中 (x_k, y_k) 为变换前某一点的坐标, (x'_k, y'_k) 为变换后对应的坐标。单一的仿射只能实现平移、缩放、旋转、翻转等操作, 为了对复杂的运动建模, 需要使用 K 组仿射变换参数分别作用于上述静态图F, 来得到 K 张变换后的图 $F' = \{f'_1, f'_2, \dots, f'_K\}$, 再通过将这 K 张变换图组合起来得到后续的一帧。假定需要生成长度为 L 的视频片段, 则共需要 $L * K * 6$ 个仿射变换参数, 本发明中将运动特征 T_m 通过一个双层的全连接网络映射为指定维度的参数。

在将 K 张变换后的图 $F' = \{f'_1, f'_2, \dots, f'_K\}$ 组合在一起的过程中, 需要评估在每个坐标位置上每张变换图的置信度, 置信度越高说明在该位置上该变换图越接近真实的视频帧, 通过引入掩码 (mask) 来实现置信度的自动评估。掩码的生成需要结合外观特征 T_a 、运动特征 T_m 和随机变量 Z, 在有初始状态以及运动趋势等信息的指导下才能更好地生成掩码。在将这些特征信息拼接后, 经过一个 3D 反卷积神经网络 (3D deconvolution neural network) 来生成掩码 $M = \{m_1, m_2, \dots, m_K\}$, 网络的最后一层使用 softmax 激活函数对每一个坐标位置 (i, j) 上的置信度进行归一化, 即满足 $\sum_{s=1}^K m_{sij} = 1$, 变换图的结合过程可以表示为式 7:

$$F_{next} = \sum_{s=1}^{s=K} m_s * f'_s \quad (\text{式 7})$$

其中 F_{next} 为生成的下一帧图像。

通过上述过程构建得到双流式视频生成器。

3) 对抗性训练过程, 通过交替训练双流式视频生成器和辨别器进行优化, 得到最优化的生成器, 使得生成的视频内容趋于真实;

本发明使用对抗性训练过程, 对抗性训练利用零和博弈的思想, 通过交替训练生成器和辨别器的方式同时提高两者的性能, 然后取用训练好的生成器来完成文本到视频的生成任务。

本发明具体采用文献[2] (Pan Y, Qiu Z, Yao T, et al. To Create What You Tell: Generating Videos from Captions[J]. 2018.) 中记载的两个辨别器: 帧辨别器和视频辨别器, 分别从每一帧的质量和整体运动的连贯性两方面来指导和训练模型, 同时加入文本特征来强化训练过程, 使得生成的视频可以符合输入文本描述, 实现生成的准确性。

执行操作如下:

31) 使用帧辨别器 (frame-level discriminator) 的方法, 帧辨别器是四层的 2D 卷积神经网络结构; 对生成的视频 $V^g = \{v_1^g, v_2^g, \dots, v_l^g\}$ 中的每一帧 v_i^g 和作为参照的真实的视频 $V^r = \{v_1^r, v_2^r, \dots, v_l^r\}$ 中的每一帧 v_i^r , 加入正负样本的文本外观特征 T_a^+ 与 T_a^- , 组成真实图文对 $R_i^v(v_i^r, T_a^+)$ 、不匹配图文对 $W_i^v(v_i^r, T_a^-)$ 以及生成图文对 $G_i^v(v_i^g, T_a^+)$ 。传统的辨别器只能缩小真实数据分布与预测分布的误差, 加入文本特征强化训练可以进一步拉近与预测的图像数据分布与真实且外观符合描述的图像数据分布之间的距离, 实现生成的准确性, 其损失函数表达式为式 8:

$$L_{D_i}^v = -\frac{1}{N} \sum_{k=1}^N c_k \cdot [\log D(R_i^v) + \frac{1}{2} \cdot \log(1 - D(W_i^v))] + \frac{1}{2} \cdot \log(1 - D(G_i^v)) \quad (\text{式 8})$$

其中, N 是一次迭代中数据的数目, c_k 为真实数据的类别, $D(R_i^v)$ 、 $D(W_i^v)$ 和 $D(G_i^v)$ 为帧辨别器对不同图文对的类别预测。帧辨别器的目标函数整体可以表示为式 9:

$$L_D^v = \frac{1}{l} \sum_{i=1}^l L_{D_i}^v \quad (\text{式 9})$$

其中 l 为生成视频的长度。

与之相对的, 上述双流式视频生成器的优化目标之一为在假定帧辨别器的预测类别为真实数据类别的前提下使得生成的图像数据分布与真实且外观匹配的图像数据分布尽可能相似, 优化函数表达式为式 10:

$$L_G^v = -\frac{1}{l} \sum_{i=1}^l \frac{1}{N} \sum_{k=1}^N c_k \cdot \log D(G_i^v) \quad (\text{式 10})$$

32) 使用视频辨别器 (video-level discriminator) 的方法, 辨别器是四层的 3D 卷积神经网络结构; 相似地, 对生成的视频 V^g 和作为参照的真实的视频 V^r , 加入正负样本的文本运动特征 T_m^+ 与 T_m^- , 组成真实视频文本对 $R^V(V^r, T_m^+)$ 、不匹配的视频文本对 $W^V(V^r, T_m^-)$ 以及生成视频文本对 $G^V(V^g, T_m^+)$ 。其作用是拉近预测的视频数据分布与真实且运动符合描述的视频数据分布之间的距离, 其损失函数表达式为式 11:

$$L_D^V = -\frac{1}{N} \sum_{k=1}^N c_k \cdot [\log D(R^V) + \frac{1}{2} \cdot \log(1 - D(W^V))] + \frac{1}{2} \cdot \log(1 - D(G^V)) \quad (\text{式 11})$$

其中 $D(R^V)$ 、 $D(W^V)$ 和 $D(G^V)$ 为辨别器对不同视频文本对的类别预测，同样的，双流式视频生成器的另一个优化目标为在假定视频辨别器的预测类别为真实数据类别的前提下使得生成的视频数据分布与真实且运动匹配的视频数据分布尽可能相似，优化函数表达式为式 12：

$$L_G^V = -\frac{1}{N} \sum_{k=1}^N c_k \cdot \log D(G^V) \quad (\text{式 12})$$

对抗训练中辨别器的目标函数由帧辨别器和视频辨别器的损失函数组成，可以表示为式 13：

$$L_D = L_D^V + L_D^F \quad (\text{式 13})$$

双流式视频生成器的目标函数同样也由两部分组成，可以表示为式 14：

$$L_G = L_G^V + L_G^F \quad (\text{式 14})$$

辨别器与生成器的目标函数交替训练，得到最优化的生成器，即可实现基于文本内容的视频生成。

与现有技术相比，本发明的有益效果是：

本发明提供了一种基于文本不同特征空间的双流式视频生成方法，根据文本中所描述的物体外观特征空间与其相应的运动特征空间，通过分离文本中所包含的空间特征和时序特征，并采用双流的方式分别对这些特征建模，利用对抗训练的方式来优化生成结果，使其逐渐趋于真实，从而实现智能化生成符合文本内容的视频片段。与现有技术相比，其技术优势体现在以下两方面：

（一）通过分离文本特征中的外观特征和动作特征，采用双流式的方式生成视频，避免了高估单一模型的能力，实现对文本特征的准确学习。

（二）通过对运动特征的单独建模，并采用一种“先静后动”的生成方式，模型自适应地可以保持时序一致性，即视频中物体运动的连贯性，而不需要如已有方法在对抗训练中添加额外的指导动作连贯性的辨别器。

附图说明

图 1 是本发明提供方法的流程框图。

图 2 是本发明实施例根据输入文本生成双流式视频的结果示例图；

其中，(a)-(j)的输入文本分别为：

（a）“The digit 0 is moving left and right.”；

（b）“The digit 1 is moving up and down.”；

- (c) “The digit 2 is moving left and right.”;
- (d) “The digit 3 is moving left and right.”;
- (e) “The digit 4 is moving up and down.”;
- (f) “The digit 5 is moving left and right.”;
- (g) “The digit 6 is moving up and down.”;
- (h) “The digit 7 is moving up and down.”;
- (i) “The digit 8 is moving left and right.”;
- (j) “The digit 9 is moving up and down.”。

具体实施方式

下面结合附图，通过实施例进一步描述本发明，但不以任何方式限制本发明的范围。

本发明提供了一种基于文本不同特征空间的双流视频生成方法，通过分离文本中所包含的空间特征和时序特征，并采用双流的方式分别对这些特征建模，最大化对指定特征的学习能力，并利用对抗训练的方式来优化生成结果。

本发明提供的方法包括：文本特征提取过程、双流式视频生成过程和对抗性训练过程；图 1 所示是本发明提供方法的流程，具体步骤如下：

1、进行文本特征提取和分离，参见步骤 11) -13)

11) 利用双向长短时记忆网络，针对输入的文本，提取正序文本特征 $T_x = \{t_{x1}, t_{x2}, \dots, t_{xn}\}$ 和逆序文本特征 $T_y = \{t_{y1}, t_{y2}, \dots, t_{yn}\}$ ， n 表示文本的长度，然后通过取均值的方式 $T = (T_x + T_y) / 2$ 得到需要的文本特征 $T = \{t_1, t_2, \dots, t_n\}$ 。

12) 引入注意力关注机制，学习两组权重 $w_a = \{w_{a1}, w_{a2}, \dots, w_{an}\}$ 和 $w_b = \{w_{b1}, w_{b2}, \dots, w_{bn}\}$ ，并且满足 $\sum_{i=1}^n w_{ai} = 1$ 和 $\sum_{i=1}^n w_{bi} = 1$ ，分别用于提取外观特征 T_a 和运动特征 T_m 。以外观特征为例， $T_a = \sum_{i=1}^n t_i * w_i$ 。权重的学习网络是两组并行的全连接网络，全连接网络的最后一层使用 softmax 激活函数。

权重中每个值 w_{ai} 代表了对文本中不同位置特征的关注程度，值越大代表其对于外观特征的描述更多，通过加权求和的方式可得到“关注”后的特征。

同理，文本特征中包含的运动特征 T_m 也可以采用这种方式来进行提取。

13) 文本特征提取和分离的训练过程，以外观特征的提取为例，通过引入三元组约束

(triplet constraint)调整文本分布,并且通过计算文本特征与图像特征在同构空间的欧式距离,即拉近匹配的文本图像对(T_a, G^+)的距离,同时增加不匹配的文本图像对(T_a, G^-)的距离,确保分离出的特征是准确的、有意义的。运动特征的提取过程与此相似,优化匹配的文本视频对(T_m, V^+)的距离,增加不匹配的文本视频对(T_m, V^-)的距离。

2、双流式视频生成过程,参见步骤 21-22)

21) 对外观特征 T_a ,先利用一个单层全连接网络结构将其映射到 $4*4*1024$ 维度,再通过 2D 反卷积神经网络(2D deconvolution neural network)来生成静态图F。2D 反卷积神经网络由 5 层反卷积层组成,每层反卷积层的感受野大小为 $4*4$,步长为 2。

为了对同样的文本生成更多多样性的视频,引入一个随机变量 Z,该随机变量从高斯分布采样,使得生成的静态图中的物体可以有略微不同的外观和不同的初始位置。

22) 对运动特征 T_m ,学习仿射变换(affine transform)的参数来对运动规律建模,通过将这些参数运用到静态图F上得到一组变换图,再与掩码结合来生成视频。掩码的生成需要结合外观特征 T_a 、运动特征 T_m 和随机变量 Z,经过一个 3D 反卷积神经网络来生成掩码,网络的最后一层使用 softmax 激活函数。

3、对抗性训练过程,参见如下步骤:

使用帧辨别器和视频辨别器,对生成的视频帧质量和视频的运动进行指导,并通过分别加入外观特征 T_a 和运动特征 T_m 强化训练过程,使生成的视频可以符合文本描述。辨别器与生成器的目标函数交替训练,得到最优化的生成器。

本发明实施主要是在文献[1](Xue T, Wu J, Bouman K L, et al. Visual Dynamics: Probabilistic Future Frame Synthesis via Cross Convolutional Networks[J]. 2016.)提供的 Moving MNIST 数据集上进行,并且通过与文献[2](Pan Y, Qiu Z, Yao T, et al. To Create What You Tell: Generating Videos from Captions[J]. 2018.)记载的方法进行比较,图 2 是本实施例进行视频生成的结果示例图,通过图 2 可以看出,本发明可以准确地生成出符合文本描述的视频片段,采用本发明模型能够自适应地生成时序一致的视频片段,不需要在对抗训练中使用额外的运动辨别器来对其指导,这也是本发明优于现有方法的技术优势。

需要注意的是,公布实施例的目的在于帮助进一步理解本发明,但是本领域的技术人员可以理解:在不脱离本发明及所附权利要求的精神和范围内,各种替换和修改都是可能的。因此,本发明不应局限于实施例所公开的内容,本发明要求保护的范围以权利要求书界定的范围为准。

权利要求书

1、一种基于文本不同特征空间的双流式视频生成方法，通过分离文本中所包含的空间特征和时序特征，采用双流的方式进行特征建模，并利用对抗训练的方式进行训练，实现智能化生成符合文本内容的真实高的视频片段；包括：文本特征提取过程、双流式视频生成器构建过程和对抗性训练过程；

1) 文本特征提取过程，执行如下操作：

11) 提取得到文本的外观特征 T_a 和运动特征 T_m ；

12) 对提取的外观特征 T_a 和运动特征 T_m ，在数据集中选择正负样本，同时引入三元组损失，准确分离文本特征，具体执行如下操作：

121) 在训练提取外观特征过程中，定义三元组 $\Phi^a = (T_a, G^+, G^-)$ ，其中： T_a 为参照用外观特征； G^+ 为正样本，是包含符合文本描述的物体的视频帧； G^- 为负样本，是包含文本描述不同的物体的视频帧；外观特征只需物体类型符合文本描述；

选取视频中的一帧；对选取的视频帧，使用 2D 卷积神经网络提取视频空间特征，并利用一个单层全连接网络结构将视频空间特征映射到 T_a 的同构空间；使用 L2 范数计算不同模态数据间的距离，即同构空间内图像与文本之间的欧式距离 $l_2^a(T_a, G^+)$ ；

以外观特征 T_a 为参照物，引入外观特征的三元组损失 $L_{tri}^a(\Phi^a)$ 调整文本分布，即拉近匹配的文本图像对 (T_a, G^+) 的距离，同时增加不匹配的文本图像对 (T_a, G^-) 的距离，确保分离出的外观特征准确；

122) 在训练提取动作特征过程中，选取包含符合文本描述的运动的视频 V^+ ；使用整个视频，可忽略视频中包含的物体类型；

对选取的视频帧，使用 3D 卷积神经网络提取视频时序特征，并利用一个单层全连接网络结构将视频时序特征映射到 T_m 的同构空间；

通过构建三元组 $\Phi^m = (T_m, V^+, V^-)$ 并采用损失函数计算运动特征的三元组损失 $L_{tri}^m(\Phi^m)$ ，拉近匹配的文本视频对 (T_m, V^+) 的距离，增加不匹配的文本视频对 (T_m, V^-) 的距离；

文本特征提取过程的目标函数 L 由外观特征的三元组损失 $L_{tri}^a(\Phi^a)$ 与运动特征的三元组损失 $L_{tri}^m(\Phi^m)$ 组成；

2) 双流式视频生成器的构建过程，执行如下操作：

21) 利用提取的外观特征 T_a ，通过 2D 反卷积神经网络生成一张静态图；

22) 利用运动特征 T_m 学习仿射变换的参数，作用于该静态图上，并通过掩码结合来衍

生出后续的视频帧；

由此构建双流式视频生成器；

3) 对抗性训练过程，通过交替训练双流式视频生成器和辨别器，并通过分别加入外观特征 T_a 和运动特征 T_m 强化训练过程进行优化，得到最优化的生成器，使得生成的视频内容符合文本描述；

通过上述步骤，实现基于文本不同特征空间的双流式视频生成。

2、如权利要求 1 所述基于文本不同特征空间的双流式视频生成方法，其特征是，步骤 11) 具体执行如下操作：

111) 对输入的文本，采用双向长短期记忆网络实现对文本特征的映射，提取正序文本特征 $T_x = \{t_{x1}, t_{x2}, \dots, t_{xn}\}$ 和逆序文本特征 $T_y = \{t_{y1}, t_{y2}, \dots, t_{yn}\}$ ，其中 n 表示文本的长度；然后通过取均值的方式 $T = (T_x + T_y) / 2$ 得到需要的文本特征 $T = \{t_1, t_2, \dots, t_n\}$ ；

112) 引入关注机制对不同类型的文本特征进行分离，得到外观特征 T_a 和运动特征 T_m ：

文本特征 T 分别通过两组全连接网络，网络最后一层使用 softmax 激活函数，分别学习两组权重 $w_a = \{w_{a1}, w_{a2}, \dots, w_{an}\}$ 和 $w_b = \{w_{b1}, w_{b2}, \dots, w_{bn}\}$ ，并且满足 $\sum_{i=1}^n w_{ai} = 1$ 和 $\sum_{i=1}^n w_{bi} = 1$ ，由此提取得到外观特征 T_a 和运动特征 T_m 。

3、如权利要求 2 所述基于文本不同特征空间的双流式视频生成方法，其特征是，步骤 111) 中所述的双向长短期记忆网络分为上下两层，分别对文本进行正序和逆序的解析；文本中的每一个词依次输入双向长短期记忆网络的每个时间节点，每个时间节点处的维度为 256 维，提取到正序文本特征和逆序文本特征。

4、如权利要求 2 所述基于文本不同特征空间的双流式视频生成方法，其特征是，步骤 112) 中，外观特征 T_a 具体通过式 2 计算提取得到：

$$T_a = \sum_{i=1}^n t_i * w_{ai} \quad (\text{式 2})$$

其中，每个权重值 w_{ai} 代表对文本中不同位置特征的关注程度。

5、如权利要求 1 所述基于文本不同特征空间的双流式视频生成方法，其特征是，步骤 121) 使用 L2 范数计算不同模态数据间距离，通过式 3 计算得到同构空间内图像与文本之间的欧式距离 $l_2^a(T_a, G^+)$ ：

$$l_2^a(T_a, M^+) = \|T_a - G^+\|_2^2 \quad (\text{式 3})$$

外观特征的三元组损失 $L_{tri}^a(\Phi^a)$ 具体通过式 4 计算得到：

$$L_{tri}^a(\Phi^a) = \max(0, \alpha - l_2^a(T_a, G^+) + l_2^a(T_a, G^-)) \quad (\text{式 } 4)$$

其中, α 为超参数, 表示文本特征与正负样本间距离之差的最大值;

运动特征的三元组损失 $L_{tri}^m(\Phi^m)$ 也采用式 4 表示的损失函数计算得到;

$$L = L_{tri}^a(\Phi^a) + L_{tri}^m(\Phi^m) \quad (\text{式 } 5)$$

文本特征提取过程的目标函数 L 表示为式 5。

6、如权利要求 1 所述基于文本不同特征空间的双流式视频生成方法, 其特征是, 步骤 21) 具体执行如下操作:

211) 利用一个单层全连接网络结构, 将外观特征 T_a 映射到 $4*4*1024$ 维度, 将维度重组为 (4, 4, 1024);

212) 再通过 2D 反卷积神经网络生成静态图; 其中, 2D 反卷积神经网络由 5 层反卷积层组成, 每层反卷积层的感受野大小为 $4*4$, 步长为 2, 输出的图像大小为 $64*64$;

213) 同时引入一个随机变量 Z , 该随机变量从高斯分布采样, 使得生成的静态图中的物体可以有略微不同的外观和不同的初始位置, 以使得对同样的文本生成更多样性的视频。

7、如权利要求 1 所述基于文本不同特征空间的双流式视频生成方法, 其特征是, 步骤 22) 利用运动特征 T_m 学习仿射变换的参数, 具体执行如下操作:

221) 单一的仿射变换包括六个参数, 表示为矩阵 $A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & 1 \end{bmatrix}$, 变换过程表示为式

6, 用于实现平移、缩放、旋转、翻转操作:

$$\begin{bmatrix} x'_k \\ y'_k \\ 1 \end{bmatrix} = A \begin{bmatrix} x_k \\ y_k \\ 1 \end{bmatrix} \quad (\text{式 } 6)$$

其中, (x_k, y_k) 为变换前某一点的坐标, (x'_k, y'_k) 为变换后对应的坐标;

222) 使用 K 组仿射变换参数分别作用于静态图, 得到 K 张变换后的图 $F' = \{f'_1, f'_2, \dots, f'_K\}$, 再通过将 K 张变换后的图组合起来, 得到后续的一帧, 以实现复杂的运动建模。

8、如权利要求 7 所述基于文本不同特征空间的双流式视频生成方法, 其特征是, 步骤 222) 中, 假定需要生成长度为 L 的视频片段, 则共需要 $L*K*6$ 个仿射变换参数, 将运动特征 T_m 通过一个双层的全连接网络映射为指定维度的参数; 在将 K 张变换后的图 $F' = \{f'_1, f'_2, \dots, f'_K\}$ 组合在一起的过程中, 评估在每个坐标位置上每张变换图的置信度; 置信度越高说明在该位置上该变换图越接近真实的视频帧;

具体通过外观特征 T_a 、运动特征 T_m 和随机变量 Z ，在有初始状态及运动趋势信息的指导下，经过一个 3D 反卷积神经网络生成掩码 $M = \{m_1, m_2, \dots, m_K\}$ ；再通过掩码实现置信度的自动评估；网络的最后一层使用 softmax 激活函数对每一个坐标位置 (i, j) 上的置信度进行归一化，即满足 $\sum_{s=1}^{s=K} m_{s_{ij}} = 1$ ；

变换图的过程表示为式 7：

$$F_{next} = \sum_{s=1}^{s=K} m_s * f'_s \quad (\text{式 7})$$

其中， F_{next} 为生成的下一帧图像。

9、如权利要求 1 所述基于文本不同特征空间的双流式视频生成方法，其特征是，步骤 3) 中，所述判别器包括帧判别器和视频判别器；对抗性训练过程具体执行如下操作：

31) 帧判别器采用四层的 2D 卷积神经网络结构；使用帧判别器的方法，对生成的视频 $V^g = \{v_1^g, v_2^g, \dots, v_l^g\}$ 中的每一帧 v_i^g 和作为参照的真实的视频 $V^r = \{v_1^r, v_2^r, \dots, v_l^r\}$ 中的每一帧 v_i^r ，加入正负样本的文本外观特征 T_a^+ 与 T_a^- ，组成真实图文对 $R_i^v(v_i^r, T_a^+)$ 、不匹配图文对 $W_i^v(v_i^r, T_a^-)$ 及生成图文对 $G_i^v(v_i^g, T_a^+)$ ，进行强化训练，由此拉近预测的图像数据分布与真实且外观符合描述的图像数据分布之间的距离，提高生成视频的准确性；损失函数表达式为式 8：

$$L_{D_i}^v = -\frac{1}{N} \sum_{k=1}^N c_k \cdot [\log D(R_i^v) + \frac{1}{2} \cdot \log(1 - D(W_i^v))] + \frac{1}{2} \cdot \log(1 - D(G_i^v)) \quad (\text{式 8})$$

其中， N 是一次迭代中数据的数目； c_k 为真实数据的类别； $D(R_i^v)$ 、 $D(W_i^v)$ 和 $D(G_i^v)$ 分别为帧判别器对不同图文对的类别预测；

帧判别器的目标函数表示为式 9：

$$L_D^v = \frac{1}{l} \sum_{i=1}^l L_{D_i}^v \quad (\text{式 9})$$

其中， l 为生成视频的长度；

双流式视频生成器的优化目标之一为：在假定帧判别器的预测类别为真实数据类别的前提下，使得生成的图像数据分布与真实且外观匹配的图像数据分布尽可能相似，优化函数表达式为式 10：

$$L_G^v = -\frac{1}{l} \sum_{i=1}^l \frac{1}{N} \sum_{k=1}^N c_k \cdot \log D(G_i^v) \quad (\text{式 10})$$

32) 视频判别器采用四层的 3D 卷积神经网络结构；使用视频判别器的方法，对生成的视频 V^g 和作为参照的真实的视频 V^r ，加入正负样本的文本运动特征 T_m^+ 与 T_m^- ，组成真实视频

文本对 $R^V(V^r, T_m^+)$ 、不匹配的视频文本对 $W^V(V^r, T_m^-)$ 及生成视频文本对 $G^V(V^g, T_m^+)$ ，进行训练，由此达到拉近预测的视频数据分布与真实且运动符合描述的视频数据分布之间的距离的效果；损失函数表达式为式 11：

$$L_D^V = -\frac{1}{N} \sum_{k=1}^N c_k \cdot [\log D(R^V) + \frac{1}{2} \cdot \log(1 - D(W^V)) + \frac{1}{2} \cdot \log(1 - D(G^V))] \quad (\text{式 11})$$

其中， $D(R^V)$ 、 $D(W^V)$ 和 $D(G^V)$ 为辨别器对不同视频文本对的类别预测；

$$L_G^V = -\frac{1}{N} \sum_{k=1}^N c_k \cdot \log D(G^V) \quad (\text{式 12})$$

双流式视频生成器的另一个优化目标为：在假定视频辨别器的预测类别为真实数据类别的前提下使得生成的视频数据分布与真实且运动匹配的视频数据分布尽可能相似，优化函数表达式为式 12。

10、如权利要求 9 所述基于文本不同特征空间的双流式视频生成方法，其特征是，对抗训练中辨别器的目标函数由帧辨别器和视频辨别器的损失函数组成，表示为式 13：

$$L_D = L_D^v + L_D^V \quad (\text{式 13})$$

双流式视频生成器的目标函数也由两部分组成，表示为式 14：

$$L_G = L_G^v + L_G^V \quad (\text{式 14})$$

通过辨别器与生成器的目标函数进行交替训练，得到最优化的生成器，即可实现基于文本内容的视频生成。

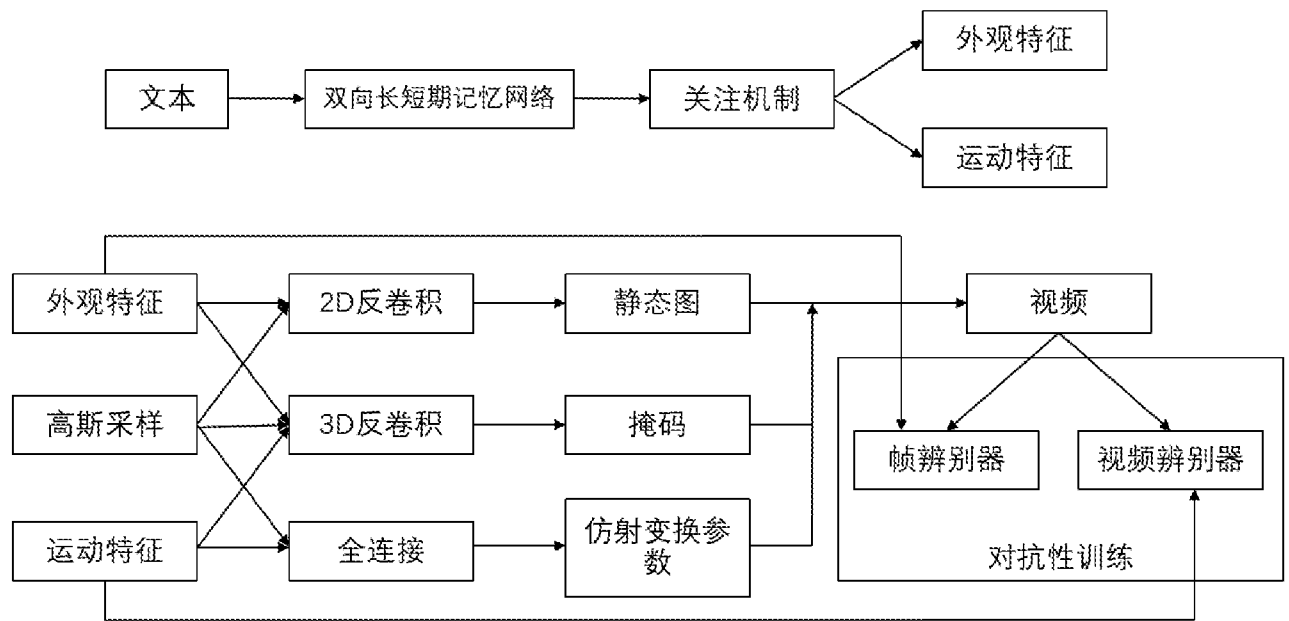


图 1



(a)



(b)



(c)



(d)



(e)



(f)



(g)



(h)



(i)



(j)

图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/085962

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2006.01)i; G06N 3/04(2006.01)i; G06N 3/08(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K;G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC, IEEE, CNKI: 视频, 生成, 文本, 特征, 时间, 时序, 空间, 时空, 外观, 运动, 静态, 动态, 图片, 帧, 神经网络, 训练, 对抗, video, generat+, text, word, character, time, space, motion, static, dynamic, picture, frame, neural net, neural network, train+, confrontation

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 108596265 A (SUN YAT-SEN UNIVERSITY) 28 September 2018 (2018-09-28) description, paragraphs [0023]-[0034]	1-10
A	CN 108090940 A (SEYYER, INC.) 29 May 2018 (2018-05-29) entire document	1-10
A	CN 108830212 A (PEKING UNIVERSITY SHENZHEN GRADUATE SCHOOL) 16 November 2018 (2018-11-16) entire document	1-10
A	CN 109325135 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 12 February 2019 (2019-02-12) entire document	1-10
A	CN 107609460 A (NANJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 19 January 2018 (2018-01-19) entire document	1-10
A	WO 2017209307 A1 (MITSUBISHI ELECTRIC CORPORATION) 07 December 2017 (2017-12-07) entire document	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

11 November 2019

Date of mailing of the international search report

28 November 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/085962

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104731959 A (BEIJING WEIYANG TECHNOLOGY CO., LTD.) 24 June 2015 (2015-06-24) entire document	1-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/085962

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108596265	A	28 September 2018	None			
CN	108090940	A	29 May 2018	US	9082400	B2	14 July 2015
				JP	6019108	B2	02 November 2016
				JP	2014519082	A	07 August 2014
				EP	2705515	A2	12 March 2014
				CN	103650002	A	19 March 2014
				CN	103650002	B	23 February 2018
				US	2013124206	A1	16 May 2013
				WO	2012154618	A2	15 November 2012
CN	108830212	A	16 November 2018	None			
CN	109325135	A	12 February 2019	None			
CN	107609460	A	19 January 2018	None			
WO	2017209307	A1	07 December 2017	DE	112017002799	T5	14 March 2019
				US	2017351935	A1	07 December 2017
				US	9971958	B2	15 May 2018
				CN	109196526	A	11 January 2019
				JP	2019510325	A	11 April 2019
CN	104731959	A	24 June 2015	CN	104731959	B	17 October 2017

国际检索报告

国际申请号

PCT/CN2019/085962

A. 主题的分类 G06K 9/62(2006.01)i; G06N 3/04(2006.01)i; G06N 3/08(2006.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																										
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G06K;G06N 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNKI, CNPAT, WPI, EPDOC, IEEE, CNKI: 视频, 生成, 文本, 特征, 时间, 时序, 空间, 时空, 外观, 运动, 静态, 动态, 图片, 帧, 神经网络, 训练, 对抗, video, generat+, text, word, character, time, space, motion, static, dynamic, picture, frame, neural net, neural network, train+, confrontation																										
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 108596265 A (中山大学) 2018年 9月 28日 (2018 - 09 - 28) 说明书第0023-0034段</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 108090940 A (西尔股份有限公司) 2018年 5月 29日 (2018 - 05 - 29) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 108830212 A (北京大学深圳研究生院) 2018年 11月 16日 (2018 - 11 - 16) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 109325135 A (平安科技深圳有限公司) 2019年 2月 12日 (2019 - 02 - 12) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 107609460 A (南京邮电大学) 2018年 1月 19日 (2018 - 01 - 19) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>WO 2017209307 A1 (MITSUBISHI ELECTRIC CORPORATION) 2017年 12月 7日 (2017 - 12 - 07) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 104731959 A (北京威扬科技有限公司) 2015年 6月 24日 (2015 - 06 - 24) 全文</td> <td>1-10</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 108596265 A (中山大学) 2018年 9月 28日 (2018 - 09 - 28) 说明书第0023-0034段	1-10	A	CN 108090940 A (西尔股份有限公司) 2018年 5月 29日 (2018 - 05 - 29) 全文	1-10	A	CN 108830212 A (北京大学深圳研究生院) 2018年 11月 16日 (2018 - 11 - 16) 全文	1-10	A	CN 109325135 A (平安科技深圳有限公司) 2019年 2月 12日 (2019 - 02 - 12) 全文	1-10	A	CN 107609460 A (南京邮电大学) 2018年 1月 19日 (2018 - 01 - 19) 全文	1-10	A	WO 2017209307 A1 (MITSUBISHI ELECTRIC CORPORATION) 2017年 12月 7日 (2017 - 12 - 07) 全文	1-10	A	CN 104731959 A (北京威扬科技有限公司) 2015年 6月 24日 (2015 - 06 - 24) 全文	1-10
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
A	CN 108596265 A (中山大学) 2018年 9月 28日 (2018 - 09 - 28) 说明书第0023-0034段	1-10																								
A	CN 108090940 A (西尔股份有限公司) 2018年 5月 29日 (2018 - 05 - 29) 全文	1-10																								
A	CN 108830212 A (北京大学深圳研究生院) 2018年 11月 16日 (2018 - 11 - 16) 全文	1-10																								
A	CN 109325135 A (平安科技深圳有限公司) 2019年 2月 12日 (2019 - 02 - 12) 全文	1-10																								
A	CN 107609460 A (南京邮电大学) 2018年 1月 19日 (2018 - 01 - 19) 全文	1-10																								
A	WO 2017209307 A1 (MITSUBISHI ELECTRIC CORPORATION) 2017年 12月 7日 (2017 - 12 - 07) 全文	1-10																								
A	CN 104731959 A (北京威扬科技有限公司) 2015年 6月 24日 (2015 - 06 - 24) 全文	1-10																								
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																										
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																										
国际检索实际完成的日期 2019年 11月 11日		国际检索报告邮寄日期 2019年 11月 28日																								
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 李晶 电话号码 86-10-53961705																								

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/085962

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	108596265	A	2018年 9月 28日	无			
CN	108090940	A	2018年 5月 29日	US	9082400	B2	2015年 7月 14日
				JP	6019108	B2	2016年 11月 2日
				JP	2014519082	A	2014年 8月 7日
				EP	2705515	A2	2014年 3月 12日
				CN	103650002	A	2014年 3月 19日
				CN	103650002	B	2018年 2月 23日
				US	2013124206	A1	2013年 5月 16日
				WO	2012154618	A2	2012年 11月 15日
CN	108830212	A	2018年 11月 16日	无			
CN	109325135	A	2019年 2月 12日	无			
CN	107609460	A	2018年 1月 19日	无			
WO	2017209307	A1	2017年 12月 7日	DE	112017002799	T5	2019年 3月 14日
				US	2017351935	A1	2017年 12月 7日
				US	9971958	B2	2018年 5月 15日
				CN	109196526	A	2019年 1月 11日
				JP	2019510325	A	2019年 4月 11日
CN	104731959	A	2015年 6月 24日	CN	104731959	B	2017年 10月 17日